DOCUMENT RESUME

ED 254 827 CS 007 949

AUTHOR Ortony, Andrew; And Others
TITLE Salience, Similes, and the Asymmetry of Similarity. Technical Report No. 332.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Academy of Education, Washington, D.C.; National Inst. of Education (ED), Washington, DC.; National Science Foundation, Washington, D.C.
PUB DATE Mar 85
CONTRACT 400-81-0030
GRANT BNS-80-04541
NOTE 52p.
PUB TYPE Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Classification; *Comparative Analysis; *Language Research; Language Usage; *Learning Theories; *Metaphors; *Models; *Reading Research
IDENTIFIERS Asymmetry; Salience; *Similes

ABSTRACT

Models of similarity have traditionally assumed that the similarity relation is symmetrical. However, when reversed, similarity statements frequently have different properties from those of the original. Previous attempts to account for the asymmetry of similarity have focused only on literal comparisons, resulting in a tendency to underestimate the degree of asymmetry in nonliteral comparisons, that is, similes. In a model of similarity proposed to account for the varying degrees of asymmetry found in all kinds of natural language comparisons, asymmetry is attributed to an imbalance in the salience of the shared attributes. Studies that have tested key aspects of the model appear to provide converging evidence for the claim that asymmetry is due at least in part to salience imbalance. (Author/HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED254827

CENTER FOR THE STUDY OF READING

Technical Report No. 332

SALIENCE, SIMILES, AND THE ASYMMETRY OF SIMILARITY

Andrew Ortony, Richard J. Vondruska,
Mark A. Foss and Lawrence E. Jones

University of Illinois at Urbana-Champaign

March 1985

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The research reported herein was supported in part by the National Science Foundation under Grant No. BNS-80-04541, in part by the National Institute of Education under Contract No. NIE-400-81-0030, and in part by a Spencer Fellowship awarded to the first author by the National Academy of Education. The second author is now with John C. Maloney & Associates, 620 North Michigan Avenue, Chicago, Illinois. Special thanks are due to Stella Vosniadou for her many valuable contributions throughout the course of this research. We would also like to thank Robert R. Hoffman for his extensive comments on an earlier draft of this paper.

CS007949

EDITORIAL BOARD

William Nagy
Editor

Harry Blanchard
Nancy Bryant
Linda Chase
Pat Chrosniak
Valerie Croll
David Dunning
Linda Fielding
Dan Foertsch
Ramona Hao
Patricia Herman
Jennifer Hughes
Barbara Hunter
Steve Levine
Margie Leys
Mary Moes
Shahin Naftchi-Ardebili
Keisuke Ohtsuka
Carol Peterman
Michael Reddix
Theresa Rogers
Marlene Schommer
Judith Scott
Ileana Seda-Santana
Janice Stewart
Behrooz Tavakoli
Terry Turner
Andrea Tyler
Lynne Webber
Ian Wilkinson

Abstract

Models of similarity have traditionally assumed that the similarity relation is symmetrical. However, when reversed, similarity statements frequently have different properties from those of the original. Previous attempts to account for the asymmetry of similarity, have focussed only on literal comparisons, resulting in a tendency to underestimate the degree of asymmetry in non-literal comparisons (i.e., similes). A model of similarity is proposed to account for the varying degrees of asymmetry found in all kinds of natural language comparisons. In this model, asymmetry is attributed to an imbalance in the salience of the shared attributes. Studies are reported which test key aspects of the model. The results appear to provide converging evidence for the claim that asymmetry of similarity is due at least in part to salience imbalance.

Salience, Similes, and the Asymmetry of Similarity

In recent years, the research devoted to modeling similarity judgments (e.g., Carroll & Arabie, 1980; Carroll & Wish, 1974; Gregson, 1975; Krantz & Tversky, 1975; Krumhansl, 1979; Nygren & Jones, 1977; Shepard, Kilpatrick & Cunningham, 1975), and the research concerned with the nature and function of metaphor (e.g., Honeck & Hoffman, 1980; Lakoff & Johnson, 1980; Miall, 1982; O'Hare, 1981; Ortony, 1979b; Sacks, 1978) have for the most part progressed independently of one another. But, similarity and metaphor are not independent problems. Certainly, it is generally acknowledged that a satisfactory analysis of the nature of metaphor will have to invoke principles of similarity. However, what is less widely recognized is that a comprehensive theory of similarity will have to account for certain facts about metaphor. That this is true can be seen by considering the fact that similes (e.g., *ballerinas are like butterflies*) are metaphorical similarity statements, which means that a theory of similarity should help explain how and why such statements differ from other kinds of comparisons. A major purpose of the present research was to demonstrate the need to extend the purview of theories of similarity to the relevant but neglected domain of similes. In the course of doing so, some limitations of existing theories become apparent. The second main purpose of the work we describe was to test some key aspects of a model proposed by Ortony (1979a) which was developed in an attempt to overcome these limitations.

Of the major contributors to theories of similarity, only Tversky (1977) even mentions the relationship between similarity and metaphor. In his model of similarity, the *contrast model*, the similarity between two entities, a and b, is expressed as a linear combination of the measures of their common and distinctive features, as shown in Equation 1.

$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \qquad (1)$$

where θ, α, and $\beta \geq 0$.

Here, s is a scale of similarity, A and B represent the feature sets of a and b respectively, and f is a measure of salience. S is a function of three disjoint sets: $A \cap B$ which is the set of features common

to both entities, A-B which is the set of distinctive features of a (i.e., those in a not found in b), and B-A which is the set of distinctive features of b. To avoid confusion in discussions about the order of the terms, we take "a is like b" to be the canonical form of (directional) similarity statements, and shall refer to the first term as the a-term, and to the second term as the b-term. Furthermore, we will use "attribute" in preference to "feature" so as to avoid any implication that we are referring to some kind of primitive or atomic constituent (e.g., Katz & Fodor, 1963). Attributes can be thought of as schema constituents (see, for example, Rumelhart & Ortony, 1975), and so can sometimes be complex and/or relational in nature. Viewed in this way, attributes can be factual or mythical, known, believed, or merely suspected--they need not be true of the entity.

The development of the contrast model (see also, Tversky & Gati, 1978; Tversky & Gati, 1982; Gati & Tversky, 1982) was motivated by evidence that judgments of similarity and dissimilarity can violate the metric axioms (see, Tversky & Krantz, 1970) upon which geometric models of similarity are based. Of particular concern to us is the fact that directional similarity judgments are not always symmetric: Atypical members of categories tend to be judged as more similar to typical members than the other way around. For example, subjects rate North Korea as more similar to Red China than they do Red China to North Korea. Tversky explains such asymmetries by arguing that in directional similarity judgments, subjects tend to focus on the a-term more than on the b-term (the *focusing hypothesis*). This is represented in the contrast model by assigning a larger value to α than to β.

The contrast model, however, is not without its problems--problems stemming from the measure of salience, f. "Salience" can be construed in at least two distinct ways. In multidimensional scaling, it refers to the extent to which an attribute is used to discriminate between entities (see, for example, Carroll & Chang, 1970). Construed in this way, the salience of an attribute has to do with how relevant that attribute is for making judgments in some domain. This seems to be the sense of "salience" employed in the contrast model, as is especially evident in Tversky's (1977) *diagnosticity principle* which maintains that the perceived similarity of pairs of

objects may change when the objects appear in different groups (i.e., when the set of objects being judged changes). According to Tversky, such changes in perceived similarity are due to changes in the diagnostic value of certain features, which implies that the salience of an attribute can change as a function of the "particular object set under study." We refer to this sense of "salience" as *relevance*. However, there is another sense of the term "salience" in which it refers, not to relevance, but to the prominence or importance of an attribute in a person's representation of an entity or category. The difference between relevance and salience can be illustrated by considering a domain like that of general anesthetics and sleep-inducing drugs. Within this category the attribute *induces sleep* has high salience for all members. But, for precisely this reason, it has little relevance to similarity or dissimilarity judgments between members because the members cannot be distinguished on the basis of this attribute. In this example, the relevance of a highly salient attribute is low. However, this is not always the case. To use an example from Tversky (1977), the attribute *real* (which for most objects has relatively low salience) can acquire considerable relevance for judgments in a domain containing both mythical and actual animals. So, relevance is a characteristic of attributes with respect to task and object domains rather than with respect to objects themselves, while salience is a characteristic of attributes with respect to objects themselves. Because we shall be concerned primarily with an unrestricted domain in which any two entities can be compared, we shall be concerned only with the salience of attributes, not with their relevance.

The distinction between salience and relevance reveals a problem for the contrast model. Since the measure, f, is defined in the model as a function of attributes without regard to their origin, one has to assume that it is intended as an object-independent measure of attribute salience. Whereas this *salience independence assumption* is appropriate for what we are calling relevance, it is not appropriate for what we are calling salience. Several considerations lead to the conclusion that the contrast model assumes salience independence, one of which is that without this assumption, the model would be incomplete because it nowhere specifies a psychologically motivated principle for assigning a salience value to an attribute which has different values with respect to a and b.

Another reason why the contrast model appears to be committed to the salience independence assumption is that salience independence is a necessary consequence of feature additivity, a characteristic which, according to Tversky, is satisfied by the function, f (Tversky, 1977, p. 332). But, now consider some object x which is comprised of a set of attributes, and suppose one adds to that set some additional attributes so as to produce a new object x'. Feature additivity entails that $f(x') > f(x)$. However, since x' is a new object relative to x, its increased salience can only be assured if the salience of the original attributes that constituted x remains the same in x' as in x. But this condition is in fact a salience independence condition. Furthermore, feature additivity and salience independence are incompatible with the implication of the *minimality axiom* that "the similarity between an object and itself is the same for all objects." (Tversky, 1977, p. 328). From the example just cited, it is clear if $f(x') > f(x)$ then $s(x', x') > s(x, x)$. Tversky's rejection of this implication is certainly consistent with the salience independence assumption.

If the independence assumption is embraced then the shared attributes (i.e., the grounds) of a comparison must be equally salient for the *a*-term and the *b*-term. One reason we believe the assumption to be invalid is that people generally consider one particular order of the terms in a comparison to be more appropriate and meaningful than the other. As indicated earlier, Tversky (1977) noted that the entity referred to by the *b*-term is generally more prototypical than the one referred to by the *a*-term. While this observation is undoubtedly correct (particularly for literal comparisons where the terms are likely to refer to members of the same category) it is not *explained* by the contrast model. We do not consider the assertion (Tversky, 1977, p. 333) that in such cases the salience of the prototype, $f(B)$, is generally greater than the salience of the variant, $f(A)$, to constitute a satisfactory explanation.

We find it more appealing to explain such asymmetries as Gildea and Glucksberg (1983) have done, in terms of the fact that speakers generally adhere to the "given/new" contract (Clark & Haviland, 1977). In similes (and indeed in all similarity statements) the "given" entity is the topic of the comparison and therefore is in the *a*-position. The "new" information that is being

communicated about the given entity is contained in the *b*-term in the sense that it is a subset of the *b*-term's attributes. Presumably, to convey the new information, a speaker selects a *b*-term for which the attributes to be communicated are highly salient. For this reason the *b*-term is likely to be a good example of something possessing those attributes. For example, suppose a speaker wants to suggest that an insult can cut deeply and cause a great deal of pain, and chooses to do so by using a simile. If the speaker's communicative intention is to be realized, it will not be sufficient to select some arbitrary *b*-term whose referent happens to have the properties of being able to cut deeply and cause pain. These attributes need to have some special status for the *b*-term if they are to be identified by the hearer as the predicates to be applied to *an insult*. The best strategy for the speaker, therefore, is to identify something for which the attributes are very salient, as is the case with *razor*. Thus, although a sheet of paper can cut and cause pain, the statement *an insult is like a sheet of paper* is virtually incomprehensible. This is presumably because in the case of *a sheet of paper*, cutting and causing pain are simply not salient enough, and the more salient properties of sheets of paper seem not to result in a coherent statement at all. This analysis not only explains why the more typical example is in the *b*-position, but it also seems to imply that the attributes to be predicated of the *a*-term are likely to be more salient with respect to the *b*-term than with respect to the *a*-term. If this were not normally the case, there would be no point in making the statement of similarity in the first place.

Apart from the theoretical reasons we have just discussed, the results of a study by Verbrugge and McCarrell (1977) and of a pilot studied we conducted, also suggest that the independence assumption is probably invalid. In their study, Verbrugge and McCarrell (1977) had subjects generate terms whose referents possessed certain attributes provided by the experimenters. Unknown to the subjects, these attributes were the grounds of metaphors. Verbrugge and McCarrell found that on average, *b*-terms from the original metaphors were twice as likely to be elicited by the grounds as were *a*-terms. In a pilot study, we also examined the likelihood of evoking the terms of similes given the shared attributes. In our study subjects were given the grounds of similes of the form "*a* is like *b*" together with a version of the similes in which one or other of the terms was

replaced with a blank. Subjects were asked to supply the missing term so as to complete the metaphorical comparison. The data showed that subjects were significantly more likely to supply the original *b*-term (or something semantically close to it) than they were to supply the original *a*-term.

However, these studies yield only indirect evidence against salience equivalence (i.e., against the independence assumption). There could be other dimensions on which *a*- and *b*-terms differ (e.g., familiarity) that would also explain these results. A more direct test is needed, especially in light of our distinction between salience and relevance. Given this distinction, the problem with the contrast model can now be reexpressed in terms of alternative interpretations of the meaning of the measure, f.

In the contrast model the measure, f, seems now to be best interpreted as a measure of relevance, whereas in the model proposed by Ortony (1979a) it was considered to be a measure of salience. This latter model, which we shall refer to as the *imbalance model*, was developed in an attempt to provide a better account of the asymmetry of judged similarity, particularly for the terms in similes. It should be noted that the imbalance model is in fact itself a contrast model. Our purpose in calling it the imbalance model is to highlight the difference in theoretical emphasis between it and Tversky's model. Although Tversky (1977, p. 328) acknowledges that "The directionality and asymmetry of similarity relations are particularly noticeable in similes and metaphors," the account of metaphor that he provides lacks the specificity of the account of similarity provided by the contrast model. Furthermore, the contrast model appears to lack the theoretical machinery to predict greater asymmetry for metaphorical than for other kinds of comparisons. In the contrast model, asymmetries in judged similarity have to be accounted for in terms of the focusing hypothesis (i.e., $\alpha \neq \beta$). This is because in any similarity judgment (directional or nondirectional) the measure of the common features, $f(A \cap B)$, is independent of the order in which the two terms are presented. The contrast model also appears to predict that the perceived similarity of the terms in literal comparisons will generally be higher than of terms from

similes. This is because, in general, one should expect the terms from literal comparisons to have more common attributes and fewer distinctive ones than the terms from similes.

In the imbalance model, the measure of the shared attributes depends only on their salience for the second term (the *b*-term), as shown in Equation 2.

$$s(a,b) = \theta f^{b}(A \cap B) - \alpha f(A - B) - \beta f(B - A) \qquad (2)$$

where θ, α, and $\beta \geq 0$.

With this formulation, a literal similarity statement can be viewed as one in which the two concepts denoted by the terms are likely to share many attributes, at least some of which are of relatively high salience for both. For example, sermons are like lectures because both are oral addresses given to groups of people, etc., and these attributes are of relatively high salience for both *sermons* and *lectures*. Therefore, the measure $f^{b}(A \cap B)$ should be high. A simile can be viewed as a similarity statement in which the shared attributes tend to be of high salience for the *b*-term but of relatively low salience for the *a*-term. Because $f^{b}(A \cap B)$ is determined by the salience of the shared attributes vis à vis the *b*-term, this measure should be relatively high compared with the level predicted by the contrast model. For example, *inducing drowsiness* is more salient with respect to *sleeping pills* than with respect to *sermons*. Thus, the contribution of the salience of that attribute to the salience of the intersection is maximized if it is based on the salience for *sleeping pills*. Finally, a statement such as *sermons are like grapefruits* (for which no consistent interpretation can readily be agreed upon) is viewed as an anomalous similarity statement because there are no relatively high salient attributes of *grapefruits* that are shared with *sermons*. In this case, the measure $f^{b}(A \cap B)$ should be low. The measure $f^{a}(A \cap B)$ is also likely to be low when metaphorical similarity statements are reversed. This is because in such statements it is unusual for there to be high salient attributes of the *b*-term (i.e., the *b*-term *after* the reversal) that are shared with the *a*-term. In other words, similes cannot be reversed without radical changes in perceived similarity and/or meaningfulness because there is an inherent imbalance in the salience of the relevant attributes shared by the two terms. In general, high salience of *b*-attributes is viewed

as a necessary condition for a meaningful (i.e., a literal or metaphorical as opposed to an anomalous) similarity statement. However, a general consequence of the model is that the asymmetry of *any* similarity statement (metaphorical or literal) can be due to differences in the salience of (at least some of) the attributes in the intersection. This account is, of course, very different from that of the contrast model where the salience of the intersection does not change as a result of a reversal.

The prediction of the imbalance model that the terms from similes will tend to be rated as highly similar seems to be at odds with the predictions of the contrast model. Proponents of the contrast model might attempt to deal with this problem by arguing that similes are understood by reinterpreting the feature space that contains the compared objects. For example, Tversky (1977, p. 349) asserts that there is "a close tie between the assessment of similarity and the interpretation of metaphors. In judgments of similarity one assumes a particular feature space ... and assesses the quality of the match .... In the interpretation of similes, one assumes a resemblance ... and searches for an interpretation of the space that would maximize the quality of the match." We find this unsatisfactory primarily because, while mathematically possible, it is cognitively implausible. It presupposes a process in which subjects first must *understand* the comparison at least to a sufficient degree to permit them to determine whether or not it is metaphorical. Then, having made this determination, a reinterpretation of the feature space is undertaken, apparently for no other purpose than to enhance the perceived similarity--a goal which, incidentally, could be more simply achieved by setting both α and β to zero. Furthermore, those attempting to explain the interpretation of similes in terms of a special simile-specific mechanism have to worry about accusations of circularity. The invocation of such a mechanism appears to presuppose that the comparison has already been interpreted, at least to the degree that it can be classified as a simile rather than as a literal or anomalous comparison. Thus, while the imbalance model suggests that asymmetry of meaning and of perceived similarity have the same underlying cause, namely salience imbalance, the contrast model offers no viable account of semantic asymmetry and no motivated basis for engaging in mental gymnastics like feature space reinterpretation designed to maximize perceived similarity. It has to divorce the process of understanding a comparison from the related process of

judging the similarity of its terms.

It would seem, then, that while both the contrast model and the imbalance model make similar predictions about literal and anomalous comparisons with respect to asymmetry, the imbalance model makes specific predictions about metaphorical comparisons that the contrast model does not. In particular, two main hypotheses can be derived from the imbalance model, but not from the contrast model: (1) *the asymmetry hypothesis:* Metaphorical similarity statements will show much greater asymmetry of similarity and meaningfulness than will either literal or anomalous similarity statements, and (2) *the salience imbalance hypothesis:* The salience of the attributes involved in metaphorical similarity statements will be much higher for the *b*-terms than for the the *a*-terms, whereas for other kinds of similarity statements this imbalance will be very much less pronounced.

In order to test these hypotheses, three experiments and a validation study were conducted. Study 1 investigated the relationship between perceived similarity, type of similarity statement, and order of terms. The main purpose of this study was to test the asymmetry hypothesis. The purpose of Study 2 was primarily to norm a pool of items from which would be selected the items to be used in Studies 3 and 4. Study 3 itself employed the salience-related measures of applicability, characteristicness, and conceptual centrality to investigate both the asymmetry hypothesis and the salience imbalance hypothesis. Finally, Study 4 tested the salience imbalance hypothesis by asking subjects to identify the attribute of each concept involved in a similarity statement that contributed most to making a comparison between them understandable. Subjects then rated the distinctiveness of these attributes relative to the concepts from which they were derived.

Study 1

The purpose of this study was to investigate the asymmetry hypothesis with three types of similarity statements, literal comparisons, metaphorical comparisons, and anomalous comparisons. In the interests of brevity and clarity, we will henceforth refer to these three types of comparisons as

Literals, Similes, and Anomalies, respectively. In the experiment, subjects were presented with both orders of the terms and they selected the order that gave rise to the more "sensible" comparison. They then rated the similarity of the terms in the statements, and classified the statements as Similes, Literals, or Anomalies. It was predicted that Similes would exhibit larger differences in perceived similarity when the order of their terms is reversed than would Literals and Anomalies.

In this and all subsequent experiments in which ratings of both orders of (especially) Similes were collected, the two orders were presented as a pair rather than randomly distributed throughout a list. This is because pilot studies had shown that when subjects are required to make similarity (or other) judgments about reversed Similes (e.g., *a magnet is like a smile*) they tend to spontaneously "re-reverse" them and end up making their judgments about the original order (*a smile is like a magnet*). This problem is significantly alleviated by having subjects make their judgments about the two orders when those orders are juxtaposed.

Despite the fact that the context in which a comparison appears often determines whether or not it is a Simile, a Literal, or an Anomaly (e.g., Ortony, Schallert, Reynolds & Antos, 1978), similarity statements were selected that were assumed to be readily classifiable by subjects without contextual support. The classification phase of this study was incorporated in order to evaluate the appropriateness of this assumption.

Method

Subjects and Materials. The subjects were 62 undergraduates who participated in partial fulfillment of the requirements for an introductory psychology course.

Eighteen Similes of the form "*a* is like *b*" were used, and from them 18 Literals were constructed by replacing the *a*-term of each of the Similes with a new term that was literally related to the *b*-term in the original comparison. Finally, 18 Anomalies were constructed by randomly pairing terms from the Literals. Any obviously meaningful comparisons that resulted from this last procedure were discarded. These materials appear in Appendix A.

For each of the 54 comparisons, the original order of the two terms was reversed and this reversed order was paired with the original order to form a comparison pair. An example of a comparison pair would be the two comparisons *an education is like a stairway* and *a stairway is like an education*. Two different randomized lists of the 54 comparison pairs were constructed. The position of each comparison in the pairs was counterbalanced. This was done systematically for each comparison type. A second, complementary version of each list was constructed in which the positions of the two comparisons in each pair were interchanged, resulting in a total of four lists.

Procedure. Subjects were run in groups of from five to fifteen persons. Each subject performed both a *Similarity Rating* task and a *Classification* task. The order in which these tasks were performed was counterbalanced. In both tasks subjects saw a list involving a different randomized order of the items. For the Similarity Rating task, subjects were instructed to decide which of the two comparisons they thought seemed more "reasonable, sensible, appropriate, etc." Subjects responded by placing a check mark alongside the comparison they preferred in each pair. They then rated the similarity of the terms in both comparisons, first rating the similarity of the terms for the order they preferred, and then for the other, nonpreferred, order. A six-point scale (1 = Not Similar to 6 = Very Similar) was provided beside each comparison, and subjects circled the number on the scale that best corresponded to their judgment.

In the Classification task, subjects also indicated which of the two comparisons in each pair they preferred. They then classified their preferred comparison as a "Regular" comparison (Literal), a "Figurative" comparison (Simile), or a "Nonsensical" comparison (Anomaly). Subjects were advised that a regular comparison is one in which the two things being compared are "really alike." They were also told that a comparison in which the two things being compared are not really alike, but which still makes sense, is a figurative comparison, and that a comparison that "makes little if any sense" is a nonsensical comparison. Subjects indicated their responses by writing the letter *R*, *F*, or *N* alongside each of their preferred comparisons.

Subjects worked through the tasks at their own pace. The average time required to complete the study was about 35 minutes.

Results and Discussion

In the analyses for this and the other studies *min F'* (Clark, 1973) was computed whenever a significant finding in either the Subjects or the Items analysis was obtained. For the sake of brevity, only significant *min F'*'s are reported. In all the studies, an attempt was made to assign an equal number of subjects to each experimental condition. In those cases where the number of subjects was unequal, the approximate method of unweighted means was employed (Winer, 1962). In addition, although in many cases the analyses of variance reported employed proportions, inspection of the distributions of the dependent measures indicated that there was no need to use normalizing transformations.

Preference data. In order to obtain a scoring criterion for the preference data, an *expected preference* for each comparison pair was pre-experimentally determined by agreement amongst the authors. In the first analysis, the proportion of matching preferences from the Similarity Rating task was used as the dependent variable. A subject's order preference was scored as a "match" if it was the same as the expected preference. An analysis of variance was performed in which the factors were Task Order (Similarity Rating before or after Classification), List (Random Order 1 or 2), Comparison Position ("*a* is like *b*" before or after "*b* is like *a*"), and Comparison Type (Literal, Simile, or Anomaly). A significant main effect for Comparison Type was obtained, $min\ F'(2, 72) = 7.85$ (for this, and all subsequent analyses, an α level of 0.05 was used as the criterion for significance). The mean proportion of matches (where .50 is assumed to indicate indifference) was higher for Similes (.86) than for Literals (.66) or Anomalies (.68).

In a second analysis, the proportion of matching order preferences from the Classification task was used as the dependent variable. The factors used in the analysis were the same as those used in the first analysis. A significant main effect for Comparison Type was again obtained, $min\ F'(2, 66) = 8.55$. The mean proportion of matches was higher for Similes (.83) than for Literals (.67) or

Anomalies (.62). This pattern was virtually the same as that found using the matches from the Similarity Rating task as the dependent variable.

Taken together, the results of these analyses demonstrate that the order of the terms in Similes tends to be much more constrained than in Literals and Anomalies. The judgments upon which these findings are based were judgments about which of the two orders of a series of comparisons was more "appropriate, reasonable, sensible, etc." These judgments pertain to the *meanings* of the statements. The results suggest that all three types of comparisons are semantically asymmetrical to some degree, but this asymmetry is particularly noticeable in Similes.

Classification data. The next analysis was performed in order to measure the degree of correspondence between subjects' classifications, and the pre-experimental classifications of the items as Literals, Similes, or Anomalies. Classification judgments were scored as "correct" if they corresponded with the pre-experimental classification. The same factors were used as in the previous analyses.

The mean proportion of correct classifications was relatively high for each Comparison Type: Literals (.79), Similes (.82), Anomalies (.75), and there was no significant main effect for this factor ($min\ F' < 1$). Although it is difficult to suggest an absolute criterion for what would constitute adequate correspondence, the absence of a main effect for Comparison Type suggests that the pre experimental classifications were not seriously biased with regard to Comparison Type.

Similarity data. In the first analysis, the similarity rating assigned to the preferred order (as determined by each subject), sim_p, was used as the dependent variable. The factors used in the analysis were the same as those used in the previous analyses. A significant main effect for Comparison Type was found, $min\ F'(2, 79) = 90.19$. For Literals, the mean sim_p was 4.89 (out of a possible 6.00, where 6 = very similar), for Similes it was 4.58, and for Anomalies it was 2.22. Scheffe' tests using the error terms from both the subjects and items analyses indicated that the mean sim_p for Similes and for Literals were significantly higher than that for Anomalies, but not significantly different from each other. These findings support the prediction of the imbalance

model that similarity ratings for Similes are generally more comparable to those for Literals than to those for Anomalies.

The asymmetry hypothesis predicts that the reduction in the perceived similarity of reversed terms will be larger for Similes than for Literals and Anomalies. An analysis testing this prediction was performed by using as the dependent variable the difference score, (sim_p - sim_{np}), based on each subject's ratings for each comparison pair. The same factors were used as in the previous analysis. A significant main effect for Comparison Type was obtained, $min\ F'(2, 135) = 32.07$. For Anomalies the mean reduction in perceived similarity was .64, for the Literals it was .72, and for the Similes it was 1.61. Scheffe' tests using the error terms from both the subjects and items analyses indicated that the reduction for Similes was significantly greater than that for Literals and for Anomalies.

These findings confirm the asymmetry hypothesis, both with respect to perceived similarity and with respect to meaningfulness. The data show that while all three types of similarity statement received reduced similarity ratings when their terms were reversed, the reduction for Similes was substantially larger than for Literals and Anomalies. The data also show that subjects found reversed Similes to be much less acceptable as "reasonable, sensible, and appropriate" statements than reversed Literals. The strong preference for the preferred order of Similes provides some evidence of the relatively high degree of semantic asymmetry of Similes.

Another important aspect of the data is the finding that the perceived similarity of the terms in the preferred order for Similes was almost the same as that for Literals. This is inconsistent with the predictions of the contrast model. As indicated in the context of our earlier discussion of Equations (1) and (2) the contrast model predicts lower similarity ratings for Similes than for Literals. One slightly surprising aspect of the data was the finding of a difference in the post hoc tests between the similarity ratings for the nonpreferred orders of Similes and Anomalies. It had been expected that reversed Similes would look very much like Anomalies. There are at least two reasonable explanations of why the ratings for reversed Similes were higher than was anticipated. One is that some of the Similes may have been interpretable in both directions, either because they

were not very metaphorical (e.g., *an artery is like a river* seems less metaphorical and less asymmetrical than *a tyrant is like a bulldozer*), or because a different meaning might be discernible from the reversed form (e.g., *a zoo is like a school*). The second possibility is that although comparisons were presented in pairs so as to dissuade spontaneous reversal, the presence of the more acceptable order may have primed the meaning thus rendering the reversed order more meaningful than it might otherwise have seemed. However, whatever the correct explanation is, it should be pointed out that if the mean ratings for the nonpreferred order of Similes (2.97) is compared with the mean values of the preferred order of Literals (4.89) and Anomalies (2.22), the reversed Similes do look more like Anomalies.

The present study highlights several phenomena that need to be explained by a theory of similarity comprehensive enough to deal with both literal and metaphorical comparisons. Since these phenomena are predicted by the imbalance model, it now becomes necessary to determine whether or not the salience of the shared attributes for the *b*-term and the *a*-term differ for Literals and for Similes in the manner predicted by the imbalance model, that is, whether or not the salience imbalance hypothesis is correct.

Study 2

Although the first experiment served its purpose, there were some problems with the items. First, for consistency, all the similarity statements from which the items were constructed used the singular form of the nouns regardless of whether or not that form was appropriate. This resulted in some rather peculiar items such as *a mountain road is like a snake*, which seems to read better in the plural. Second, it became apparent that some of the Similes employed in Study 1 were either not very metaphorical and/or not widely understandable. For these reasons, Study 2 was designed to validate the materials to be used in the next two experiments. In this validation study, most of the Similes used plural rather than singular nouns except in cases where the singular clearly sounded better. Some of the Similes used in the earlier experiments were changed so as to improve them, and

for each comparison an abridged version of the shared attributes (henceforth called the *grounds*) was constructed. One of the main goals was to determine whether or not the grounds that we had constructed adequately captured the basic meaning of their corresponding comparisons. In addition, it provided an opportunity to measure the degree to which each statement was judged to be metaphorical.

Method

Subjects and Materials. Twelve undergraduates participated in partial fulfillment of a requirement for an introductory psychology course.

One hundred and twenty comparisons of the form "*a*(s) is/are like *b*(s)" were used as stimuli. Sixty of these comparisons were considered to be Similes, and 60 were considered to be Literals. The Literals were formed by replacing a term in the 60 Similes with another term that was literally related. Half of the Literals were formed by replacing an *a*-term and half were formed by replacing a *b*-term. This being the case, the order of the terms in Literals did not always correspond to an expected preferred order. These comparisons appear in Appendix B.

A brief statement of the grounds was prepared for each comparison (see Appendix B). For both the Similes and the Literals, an attempt was made to construct the best possible grounds for 48 of the 60 comparisons. For the remaining comparisons (the control Literals and Similes), an attempt was made to construct grounds that failed to capture the basic meaning, but which still were applicable to the terms. The control comparisons were included to encourage full scale usage. Two randomized lists of 60 comparisons each were constructed. The comparisons were divided between the lists in such a way that if a given Simile appeared in List 1, the Literal derived from it appeared in List 2. This insured that there would be no repetition of terms in either list.

Subjects were provided with a four-point response scale and were asked four questions about each comparison. The first question, the Agreement question, asked subjects to rate the extent to which they agreed with the comparison (1 = Strongly Disagree to 4 = Strongly Agree). The Agreement question was included mainly to distinguish agreement from adequacy of the grounds.

The second question, the Metaphoricity question, asked subjects to rate the extent to which they thought the comparison was metaphorical (1 = Not At All to 4 = Very Much). The third question, the Adequacy of Grounds question, asked subjects to evaluate how adequately the grounds presented along with the comparison captured its basic meaning (1 = Captures the Meaning Very Inadequately to 4 = Captures the Meaning Very Adequately). The subjects were instructed to make their evaluation of adequacy without regard to whether or not they agreed with the comparison. Finally, and less important, subjects were given a Correspondence question which asked them to rate the extent to which the grounds corresponded to the explanation of the meaning of the comparison that they themselves would have provided.

Procedure. Before beginning the study, subjects were presented with the comparison *arteries are like rivers*. Each of the four questions for this example was accompanied by a parenthetical statement to clarify the task. Subjects then received a booklet containing either the comparisons from List 1 or those from List 2. They were instructed to work through the booklet at a steady pace, and to respond to items in the order in which they appeared in the booklet. After completing that booklet, subjects received a second booklet containing the items from the other List. The order in which subjects received the two booklets was counterbalanced. The average time needed to complete the entire study was about 45 minutes.

Results and Discussion

The mean ratings and standard deviations for the four questions appear in Table 1.

Insert Table 1 About Here

Since our primary concern was that the *individual items* had the appropriate characteristics for inclusion in the next study, these aggregate data are provided merely to show the trends for each question. Overall, subjects agreed with Literals more strongly than they did with Similes. Because Similes tend to reflect opinions (e.g., *salesmen are like bulldozers)* rather than facts (e.g., *steamrollers*

are like bulldozers), this finding is understandable. However, on balance, the agreement ratings for the Similes were on the agreement side of the agree/disagree midpoint of 2.5. As expected, Similes received higher metaphoricity ratings than Literals, with the two types having non-overlapping distributions. The Adequacy of Grounds ratings indicated that the grounds did capture the basic meaning of each comparison. Moreover, high ratings on the Correspondence question suggested that the subjects believed that they could not have done much better in constructing the grounds themselves. Inspection of the ratings for individual items indicated that the variance was low enough to enable the adoption of reasonable criteria of adequacy and metaphoricity for the inclusion of items in the next experiment.

Study 3

The purpose of this study was to test the salience imbalance hypothesis by using three different but related measures that were intended to converge upon salience. These measures were assumed to share substantial common variance, but each was also assumed to have unique variance. In an earlier pilot experiment, an attempt had been made to test for salience imbalance using the single measure of *importance*. It was clear from the data that subjects did not have a uniform criterion for judging the importance of an attribute with respect to some object. A question such as "How important is *inducing drowsiness* for lectures?" appears to be difficult to answer because it seems to lead subjects to respond "Important with respect to what?". As a precaution against a similar problem arising with a direct inquiry about salience, three relatively specific measures were employed in the present experiment.

To the extent that an attribute is salient with respect to a concept (at least when the concept under consideration is a general term), that attribute can be expected to apply to all or virtually all instances of the concept. However, although salient attributes of a concept apply to most of its instances, it is not the case that all widely applicable attributes are highly salient. For example, the attribute *has a windshield* applies to all (or virtually all) cases of automobiles but does not seem to

be a very salient property of them. Thus the *applicability* measure is a reasonable, but imperfect, measure of salience. Another measure of salience, with a different set of imperfections, can be achieved by using judgments of the conceptual centrality of an attribute. The attribute *induces drowsiness*, for example, can be viewed as conceptually central to sleeping pills. Thus, *conceptual centrality* was used as a second measure of salience. Attributes that are conceptually central to an entity, however, do not necessarily exhaust the attributes that are salient. Some attributes might be very typical or characteristic and yet not be conceptually central. For example, the attribute *being hairy* is very characteristic of gorillas, but this attribute does not seem to be conceptually central. A shaven gorilla is no less a gorilla for that! Few people would deny, however, that *being hairy* is a salient attribute of gorillas. *Characteristicness* was used as a third measure in this attempt to converge on salience.

Method

Subjects and Materials. Fifty-three undergraduates participated in partial fulfillment of the requirements for an introductory psychology course.

The first step in preparing materials was the selection of 40 Similes and their corresponding grounds on the basis of the information obtained from the previous study. Of the 48 Similes with adequate grounds, the 40 items selected all had ratings on the Metaphoricity question of 2.50 or higher, and their corresponding grounds had Adequacy ratings of 3.00 or better. The 40 Literals originally constructed from those 40 Similes, along with their corresponding grounds, were also selected. All of these Literals had ratings on the Metaphoricity question of less than 2.00, and their corresponding grounds had Adequacy ratings of 3.00 or better. In order to encourage full scale usage, the 12 control Similes and their 12 associated control Literals, all having inadequate grounds, were also included, resulting in a total of 104 items.

The stimuli for the study were constructed by isolating the constituent *a*-terms and *b*-terms from each Simile and Literal. These terms were then paired with the corresponding grounds, resulting in 208 stimuli. The verb of each of the grounds was modified (e.g., from *induce*

drowsiness to *inducing drowsiness*). Two lists of 104 stimuli each were constructed. In each list, the grounds for a particular item were used twice--once in conjunction with the *a*-term, and once in conjunction with the *b*-term. Half of the items in each list were constructed using Similes, and half using Literals. If a particular Simile was used in List 1, its associated Literal was used in List 2, and vice versa. This was done to prevent the repetition of individual terms in each of the lists.

Procedure. Subjects were randomly assigned to one of three groups. Within each group, half of the subjects received List 1, and half received List 2. In one group subjects were asked the Applicability question. For example, they might be asked: "*Being ugly and growing uncontrollably* is believed to apply to *slums* in: 1 = Almost No Instances, 2 = Some Instances, 3 = Most Instances, or 4 = Virtually All Instances." In the second group, subjects were asked the Centrality question, an example of which was "How central to the concept of *tumors* is *being ugly and growing uncontrollably*? 1 = Not At All to 4 = Very Much." In the third group, subjects were asked the Characteristicness Question. For example, they might have been asked "How characteristic of *illiteracy* is *inhibiting freedom and restricting opportunity*? 1 = Not At All to 4 = Very Much."

After completing their task, subjects in all three groups were provided with another booklet containing 52 comparisons that were reassembled from the terms they had just encountered. Each comparison was presented along with its reversed order (as in Study 1), and the order of presentation was counterbalanced so that the original comparison appeared first for half of the pairs, and second for the other half. Within each of the three groups, half of the subjects received the pairs prefaced by the phrase "With respect to (grounds)", and half received the pairs only. This manipulation was included because it was considered possible that the reversal of the terms in a Simile might lead subjects to try to interpret the comparison using completely different grounds. For example, a Simile such as *surgeons are like butchers* would presumably be interpreted using grounds dealing with lack of finesse, whereas its reversed form, *butchers are like surgeons*, would presumably be interpreted using grounds dealing with manual dexterity at dissection. Subjects were instructed to read both orders of each comparison, and to indicate which of the two they felt was

more "reasonable, sensible, appropriate, etc." They were also asked to indicate whether or not the comparison they preferred was "meaningful" or "not meaningful", and to then do the same for the other, nonpreferred comparison. Throughout the experiment, subjects worked at their own pace.

Results and Discussion

Rating data. The ratings from each of the measures were averaged across subjects and used as the dependent variable in an analysis of variance in which the factors were Source of Item (Literal or Simile), Question (Applicability, Characteristicness, or Centrality), Term (*a*-term or *b*-term), and List (List 1 or List 2). A significant main effect was obtained for Source of Item, *min* $F'(1, 105) = 43.13$. Across all questions, the mean rating for terms from for Literals (3.26 out of a possible 4.00) was higher than that for terms from Similes (2.74). Also, a significant main effect for Term was obtained, *min* $F'(1, 88) = 8.12$. More important, the Source of Item by Term interaction was significant, *min* $F'(1, 101) = 13.56$. This interaction, in which the ratings of *a*-terms relative to *b*-terms was lower for Similes than for Literals, is predicted by the salience imbalance hypothesis. The main effect for Question was also significant, *min* $F'(2, 65) = 3.97$. The overall ratings were highest for Characteristicness (3.16), next highest for Applicability (2.95), and lowest for Centrality (2.89). However, the Question factor did not enter into significant interactions with any of the other variables.

The main prediction of the salience imbalance hypothesis is clearly confirmed by these findings. The salience of the grounds, using the three measures individually and using a composite of the three, was higher for the *b*-terms of Similes than for their *a*-terms. Since the Literals were not constructed in a preferred order, it was necessary to utilize subjects' preference data in order to investigate salience imbalance in literal comparisons.

Preference data. The first step in analyzing the preference data was to determine the proportion of subjects selecting the *a*/*b* order as their preferred order for both Similes and Literals. It was found that for Similes this proportion was .85 when the grounds were not present, and .81 when they were, again showing, as in Study 1, that Similes are not readily reversible. The

equivalent proportions for the Literals was .54 in both cases. This proximity to indifference was expected because, as mentioned in the Method Section of Study 3, the way in which the materials were constructed resulted in a more or less random assignment of terms to the *a* and *b* positions.

It is already clear from this Study, as well as from Study 1, that Similes exhibit a high degree of both asymmetry and salience imbalance. The next analysis was undertaken to determine whether this relationship also occurred for Literals. To this end, a Pearson correlation was computed between a measure of asymmetry and a measure of imbalance. The proportion of subjects selecting one order of a comparison in preference to the other was taken as the measure of asymmetry, and the difference between the means of the salience-related ratings for the *b*-terms and for the *a*-terms was taken as the measure of salience imbalance. When no grounds were presented, a significant relationship was found between the two measures, $r = .39$, $df = 38$. When the salience imbalance measure was positive, subjects showed a preference for the *a*/*b* order. When the salience imbalance measure was negative, they showed a preference for the *b*/*a* order. For example, subjects rated the grounds *wearing uniforms and being members of the armed forces* as being much more salient with respect to *soldiers* than with respect to *sailors*, and they also strongly preferred *sailors are like soldiers* (*a*/*b*) to *soldiers are like sailors* (*b*/*a*). Examination of the scatter plot indicated the existence of a few highly discrepant items. Inspection of the grounds for these items revealed plausible reasons for these discrepancies. For example, subjects strongly preferred *switchblades are like daggers* (*a*/*b*) over *daggers are like switchblades* (*b*/*a*). Unfortunately, the grounds *being sharp weapons that are easily concealed* included the phrase *easily concealed*, which seems (in retrospect) more applicable to *switchblades* than to *daggers*. Thus, the complete grounds were rated as more salient for the *a*-term than for the *b*-term--the exact opposite of the pattern that was expected for items where the *a*/*b* order was preferred.

A review of the grounds (see Appendix B) indicated that for seven items there seemed to be a bias in the phrasing of the grounds. Sometimes the grounds appeared to favor the *a*-term (*intermissions/time-outs*, *switchblades/daggers*, and *steamrollers/bulldozers*) and sometimes they

seemed to favor the *b*-term (*visa/passport, farms/zoos, clouds/mist,* and *reform school/prison*). When these items were deleted, the correlation between asymmetry and imbalance increased to $r = .80$, $df = 31$.

When the grounds were presented during the preference judgments, the strength of the relationship between imbalance and asymmetry in Literals was attenuated. For the complete set of items the correlation was not significant ($r = .28$, $df = 38$), but it was significant ($r = .46$, $df = 31$) when the seven items with biased grounds were excluded. Presumably, when the grounds are presented, subjects attempt to extract an interpretation from each of the two orders of the comparison and then select the one which yields the more sensible interpretation. Should these two interpretations be the same, the order that expresses this interpretation more appropriately will be selected. Insofar as the grounds are viewed as being a viable interpretation for both orders, only the latter consideration would enter into a preference judgment. Therefore, presentation of the grounds would be expected to reduce the degree to which one order is preferred over the other.

The finding that when Literals do show imbalance, the direction of this imbalance tends to correspond to the order of the terms that is preferred (i.e., higher salience for the term that is preferred in the *b*-position) has important implications for theories of similarity in general because it suggests that salience imbalance contributes to asymmetry even in Literals. The contrast model accounts for such differences solely in terms of the weights α and β, an explanation which the present data suggest may be incomplete.

Meaningfulness data. The last analysis was performed on the binary "meaningfulness" judgments that subjects made about both orders of the comparisons in the second part of the experiment. An analysis of variance was performed using as the dependent variable the proportion of subjects assigning a "Meaningful" response to each comparison. The factors were Comparison Type (Simile or Literal), Term Order ("*a* is like *b*" or "*b* is like *a*"), and Grounds Presentation (Present or Absent). Because in this analysis Comparisons served as the replication factor, standard *F*-values were derived. A main effect for Ground Presentation was obtained, $F(1, 78) = 6.2$. When

the grounds were presented, the proportion of comparisons classified as meaningful was .73, and when they were not, the proportion was .69. This factor was not, however, involved in any significant interactions. Main effects for Comparison Type, $F(1, 78) = 43.88$, and Term Order, $F(1, 78) = 142.70$, were also significant. But the significant interaction of these two factors, $F(1, 78) = 131.22$, was of greatest interest. Although the mean proportions of meaningful responses for the *a/b* Literals (.80) and for the *b/a* Literals (.79) were virtually identical, there was a large difference between the proportions for the *a/b* Similes (.82) and for the *b/a* Similes (.44). This finding supports the earlier claim (see Study 1) that Similes are *semantically* asymmetrical, since their meaningfulness is not preserved by the reversal of the constituent terms.

Study 4

Although Study 3 lends strong support to the salience imbalance hypothesis, it does so by examining the salience of experimenter-provided grounds. But, as was evident from the analysis of salience imbalance in Literals, measures of salience appear to be very sensitive to the particular phrasing chosen for the grounds. If salience imbalance can be demonstrated when subjects provide their own grounds then the (implicit) assumption that the experimenter-provided grounds correspond to those that make the similarity statements understandable to subjects would no longer have to be made. Pilot experiments, however, showed that subjects often find it difficult, if not impossible, to articulate the grounds of (especially metaphorical) comparisons. The purpose of the present experiment was to examine the salience imbalance hypothesis in a manner that would avoid experimenter specification of the grounds while still not requiring subjects to articulate the grounds. This compromise was achieved by requiring subjects to generate, not the grounds, but only the one attribute of each term that they thought most contributed to making the corresponding comparison understandable. This approach does not depend on the strong assumption that there must be an identity match between the shared attributes of the two entities, but rather upon the weaker assumption that such shared attributes as there are need only be highly similar to one another (Ortony, 1979a).

Method

Subjects and Materials. Thirty-six undergraduates participated in partial fulfillment of a requirement for an introductory psychology course.

Two sets of 16 comparisons were drawn from those used in the previous study. In each of the sets, eight were Similes and eight were Literals. The Literals selected were totally unrelated to the Similes (i.e., they contained none of the same terms). These materials appear in Appendix C. Within each set, the order of the comparisons was randomized, and either the *a*-term or the *b*-term was designated as the *target term.*

Procedure. Subjects were run in groups ranging from two to eight persons, and they received either Set 1 or Set 2 of the materials. They were told that they would hear a series of comparisons, and that they should write each comparison onto one of the index cards provided. They were further instructed that either the first or the second term of the comparison would be repeated, and that they should underline this (target) term. Half of the time the first term was the target term, and half of the time the second term was the target term.

After each item was read, subjects determined which aspect (attribute) of the concept denoted by the target term contributed most to making the comparison understandable. As a measure of salience, subjects were then asked to rate the distinctiveness of the attribute on a five-point scale (1 = Not At All to 5 = Extremely). A distinctive attribute was characterized as an aspect of an object that is highly prominent (i.e., easily brought to mind), important (i.e., very central or characteristic), and discriminative (i.e., distinguishes the object from other objects). Two practice trials were conducted before the main part of the study commenced. Subjects were given 30 seconds to write down each attribute and to rate its distinctiveness.

After subjects had been read all of the 16 comparisons in a set, they heard the entire set once again. On this second pass through the set, subjects again wrote down each comparison on an index card, but this time the other term in each comparison was designated as the target term. In this way, the attributes of both of the concepts involved in each comparison that contributed most to

making the comparison understandable were obtained from every subject.

Having completed this phase of the experiment, subjects were asked to reread each comparison and both of the attributes that they had generated for it. They then rated the similarity in meaning of these two attributes using a five-point rating scale (1 = Not At All to 5 = Extremely). These data were collected in order to determine whether there would be any systematic relationship between the degree of salience imbalance and the similarity of the attributes. Furthermore, it was felt that if the attributes were judged as being similar to one another, then it was more likely that both would be related to the same element in the intersection set. Subjects performed this task at their own pace, and wrote the rating on the back of each of the index cards.

Results and Discussion

Inspection of the data indicated that in virtually all cases subjects generated attributes that were relevant to the meaning of the comparison. In fact, compared with the other experimental tasks reported, this one seemed to facilitate comprehension of the comparisons to the greatest degree.

Using the salience ratings as the dependent variable, an analysis of variance was performed in which the factors were Set (Set 1 or Set 2), Comparison Type (Simile or Literal), and Term (*a*-term or *b*-term). A significant main effect for Comparison Type was obtained, $min\ F'(1, 456) = 10.22$. The mean salience ratings for the attributes generated when the target terms were in Literals (4.15) was higher when the target terms were in Similes (3.66). More important, there was a significant Comparison Type by Term interaction, $min\ F'(1, 60) = 5.24$, as shown in Table 2.

Insert Table 2 About Here

Scheffé tests indicated that the difference between the mean salience ratings for the attributes generated for the *b*-terms and *a*-terms in Similes (.29) was significant but that the difference for Literals (-.11) was not. Since the Literals employed in the present experiment were a subset of those employed in Study 3, we conclude that the reverse imbalance trend occurred for the same reason. In

any event, the finding of a significant difference in the predicted direction for the Similes again supports the salience imbalance hypothesis.

Analyses of the ratings of the "similarity in meaning" of the attributes for both terms did not yield any significant difference. In particular, there was no evidence of any relationship between the rated similarity of meaning of the attributes of the *a*-terms and *b*-terms and the asymmetry of the corresponding comparison. However, it is interesting to note that the mean similarity rating assigned to the attributes in Literals (3.77) was virtually the same as that of the attributes in Similes (3.73) and that both were relatively high. The comparability of these ratings, together with their relatively high level, lends credence to the supposition that the attributes in both Similes and Literals were indeed components of the grounds. In other words, they were shared attributes in the sense that they were similar to one another. This implication serves to increase confidence in the conclusion that the present experiment constituted a satisfactory compromise between the pre-experimental preparation of the grounds on the one hand, and the requirement that subjects generate the grounds on the other.

General Discussion

Two main hypotheses were tested and confirmed by the experiments we have reported. The first was the asymmetry hypothesis which predicts that the terms in similes should show a greater asymmetry in judged similarity than the terms in other kinds of comparisons. The hypothesis also predicts that the meaning of similes will be more affected by term reversal. With respect to judged similarity, the hypothesis was supported in Study 1 by the finding of a larger difference in the rated similarity of terms between the preferred and nonpreferred order of metaphorical comparisons than of literal and anomalous comparisons. The imbalance model predicts such a difference because it allows the measure of the intersection to be affected by the order of the terms, whereas the contrast model does not. With respect to meaningfulness, the asymmetry hypothesis was supported by findings in Studies 1 and 3 in which subjects showed stronger preferences for one of the two orders

of similes than they did for other kinds of comparisons. Again, the imbalance model accounts for this finding in terms of the effect of term order on the intersection set. In any comparison, salient attributes of the *b*-term that are shared with the *a*-term are largely responsible for the *meaning* of the statement. In reversed similes there usually are no such attributes, so they tend to be perceived as having little or no meaning. This is not the case for the "correct" order of similes, or for either order of typical literal comparisons.

The second main hypothesis, the salience imbalance hypothesis, predicts that the salience of the shared attributes in similes will be lower for *a*-terms than for *b*-terms, but that a high degree of imbalance in this direction will not be a characteristic of other kinds of comparisons. This hypothesis was supported by the findings of Studies 3, and 4. In these studies, *b*-terms received higher ratings for shared attributes than *a*-terms, regardless of the salience-related measure employed. This finding also suggests that the salience independence assumption of the contrast model is untenable.

The present findings bear on a number of corollary predictions of the imbalance model and the contrast model. In discussing equations (1) and (2) in the Introduction, it was argued that the contrast model predicts that the rated similarity of the terms of metaphorical comparisons should be lower than that typically found for literal similarity statements. This is because, for metaphorical comparisons, the measure of salience for the intersection set is likely to be relatively small compared to that for the distinctive sets, whereas for literal comparisons it is likely to be relatively large. In fact, however, Study 1 revealed no difference between the rated similarity of terms in metaphorical and literal comparisons. We believe that this result poses something of a problem for the contrast model. Unless one assumes that the salience of the matching attributes depends on the entity with which they are associated, and that in the case of similes the salience of these attributes is relatively high (deriving from the values for the *b*-term), there would seem to be no way of accounting either for the high similarity ratings accorded to metaphorical comparisons, or for the large reduction resulting from term reversal. On the other hand, the comparability of similarity ratings for the

terms in literal and metaphorical comparisons is compatible with the imbalance model. This is because the model permits the upper bound on the measure of salience of the intersection set to be increased, and thus is not committed to the prediction that the rated similarity of terms from similes will be lower than for those from literal comparisons.

The finding in Study 1 that the reduction in similarity ratings for reversed metaphorical comparisons was to a level comparable to that for (the preferred order of) anomalous comparisons also provides compelling evidence in favor of the imbalance model. According to the imbalance model, except in the few cases in which a (different) meaning can be assigned to reversed similes, there are no high-salient attributes of *b*-terms in such comparisons that are part of the intersection set. Consequently, the measure $f^{b}(A \cap B)$ is likely to be quite small, and in principle, could be comparable in size to that found for anomalous comparisons. In the contrast model, because the measure $f(A \cap B)$ is unaffected by the order of terms, there would appear to be no principled way in which to account for the low similarity ratings assigned to the terms from reversed similes, given the high ratings observed when the terms were presented in the preferred order.

Other researchers investigating metaphor have obtained results that are compatible with those reported here. For example, Harwood and Verbrugge (1977) found that the perceived similarity tended to be lower, although not significantly so, when the terms from metaphorical comparisons were presented in the reversed order--a result consistent with our asymmetry hypothesis. Since, for any item, only one order of the terms was presented to a subject in their experiment, spontaneous reversal (cf. Study 1) may have attenuated the difference between the reversed and non-reversed orders. The salience imbalance hypothesis was supported in a series of studies conducted by Hanson (1982). Hanson found that metaphorical comparisons tended to display greater salience imbalance than literal comparisons, although literal comparisons involving prototypes and variants also displayed moderate imbalance. Support for the imbalance model is also reported by Katz (1982), who specifically set out to test it. Using a variety of salience-related measures (dominance, typicality, fluency, and imaginal distinctiveness), Katz found that metaphorical statements having

shared features (attributes) that were low salient for the topic (*a*-term) and high salient for the vehicle (*b*-term) were rated as better metaphors than statements that had the opposite pattern of imbalance (i.e., high-low).

A general problem relating to the notion of salience is that it appears to suffer from the same lack of defining properties as do many other concepts (cf., Rosch & Mervis, 1975; Smith & Medin, 1981). We have tried to characterize salience in terms of applicability, conceptual centrality, characteristicness, importance, prominence, and distinctiveness. Although the data resulting from these various characterizations seem to produce the same pattern of results, we suspect that these indices still fail to capture an important aspect of salience. All seem to relate to the degree to which a term represents a good example of something possessing the attributes. But it also seems plausible that some attributes are salient with respect to particular objects only because they are conventionally symbolized by those objects. For example, the attribute *being proud* is conventionally symbolized by *peacocks*, and is a salient property of peacocks not because it is conceptually central or characteristic, or even true, but rather by virtue of its conventional symbolic association with peacocks. Perhaps one has to conclude that the most universal property of a salient attribute is that can be brought readily to mind, and that variables like characteristicness, conceptual centrality, and conventional symbolic value all contribute to the ease with which this occurs in the context of some particular word. This is an appealing notion since, ultimately, the point about salience is that more salient properties are presumed to be more readily "accessible" than less salient properties.

A close relation between salience and accessibility is consistent with Barsalou's (1982) distinction between *context independent* and *context dependent* properties of concepts (see also Cohen's, 1979 distinction between *semantical* and *practical* features). According to Barsalou, context independent properties "form the core meanings of words," and are activated by the word in all contexts. Context dependent properties, on the other hand, are activated in only some contexts. Barsalou proposed that many metaphorical comparisons may involve properties that are context

independent (and therefore highly salient) for the *b*-term and context dependent (and therefore less salient) for the *a*-term. Literal comparisons typically involve only context independent properties (which are highly salient for both terms). One interesting feature of this proposal lies in its identification of another perspective on the notion of salience, even though we have reservations about a categorical distinction between the two types of properties, partly because of the evidence from Study 3 that literal comparisons also exhibit salience imbalance.

If salience is construed in terms of ease of "accessibility," and if the most salient attributes of a concept are accessed in all or virtually all contexts, then one would expect highly salient attributes to be activated even in cases where they are unnecessary, inappropriate, or even counter-productive with respect to the task at hand. This suggests, as Barsolou (1982) noted, that subjects might be unable to "block" the metaphorical potential of statements, even when they are instructed to interpret them literally, as Glucksberg, Gildea, and Bookin (1982) found.

We believe that our results demonstrate that the imbalance model is sufficiently powerful to account for similarity data derived from metaphorical as well as literal comparisons. Furthermore, it predicts a (detected) source of asymmetry even in literal comparisons that is beyond the theoretical scope of the contrast model. However, both the imbalance model and the contrast model suffer from limitations. One of these is that both models are incomplete. As discussed in the introduction, the measure, f, employed in the imbalance model focuses on an attribute's salience, while Tversky's presentation of the contrast model focuses on relevance. Since both are pertinent aspects of an attribute, a comprehensive theory of similarity needs to incorporate them both. This would require elements of both the contrast model and the imbalance model. Assuming X_i to be some shared attribute, and h to be a measure of salience, such a combined model would have to accommodate the fact that the salience of X_i with respect to entity a, $h(X_i \mid a)$, could be different from its salience relative to entity b, $h(X_i \mid b)$. Therefore, the measure of salience of attribute, X_i, in the context of some particular task would have to be a function of these two values, for example:

$$h(X_i) = w_a h(X_i \mid a) + w_b h(X_i \mid b) \qquad (3)$$

where $w_a + w_b = 1$.

In directional similarity tasks, such as those addressed by the imbalance model, the weights are assumed to be $w_a = 0$ and $w_b = 1$. However, in nondirectional tasks it may be reasonable to assume that $w_a = w_b$. It would further have to be assumed that some other function, say $g(X_i)$, provided a measure of the relevance of X_i in the context, and that the overall measure of the intersection (and of the distinctive sets for that matter) was obtained from some additive function over the products of $g(X_i)$ and $h(X_i)$ for all values of i. In fact, it is interesting to note that one of the measures employed by Tversky (1977, p.338) appears to involve both salience and relevance. However, this measure of salience for the common attributes, is a product function, $h(X_i) = h(X_i \mid a) \cdot h(X_i \mid b)$, as opposed to the additive function proposed in Equation 3. As an estimate of attribute salience relative to each specific entity, $h(X_i \mid a)$ and $h(X_i \mid b)$, Tversky used the proportion of subjects who indicated that the attribute was characteristic of the entity. Although this measure is an improvement over ignoring entity-specific salience (as he did in the immediately preceding estimate of f), it suggests that as the salience of X_i for a decreases, so should the resulting similarity. Our proposed measure, on the other hand, does not suggest such a change in similarity. As a measure of relevance, Tversky's employed the proportion of entities in the domain that subjects indicated possessed the attribute, but for some inexplicable reason, only applied it to the measure of the common attributes. It would seem that a great deal of further theoretical and empirical work is needed in order to arrive at a satisfactory model that embodies both salience and relevance.

Another limitation of both models lies in their emphasis on the contribution of the intersection set to judgments of similarity. By doing this, both seem to exclude the possibility that a similarity statement can ever convey information that is genuinely new in the sense of previously not being a part of the representation of the a-term. This is hardly reasonable. If a child hears an adult talking about tangelos and asks what they are, it seems implausible to suppose that the child would learn nothing by being told that tangelos are like oranges.

Now insofar as the models account for perceived similarity, this may not present a problem. The problem arises if one wishes to extend them to account for the processes involved in making similarity judgments. It would seem that the imbalance model might have an advantage over the contrast model in this respect. Additional assumptions can easily be incorporated to try to account for the comprehension process. One of these assumptions (discussed in Ortony, 1979a) is that for attributes to be considered as shared, they need only be highly similar (rather than identical), as was observed in Study 4. Another additional assumption that might be helpful in developing a process model would be to incorporate a salience criterion below which attributes associated with the *b*-term would not be considered for attempted application to the *a*-term, unless primed by the context (see Gildea & Glucksberg, 1983). This could reduce the need for extensive searching for applicable attributes, although such searching would still be possible if the criterial level were lowered, as might happen when one attempts to make sense of a simile whose meaning is not immediately apparent. One might also have to postulate a lowering, or even the elimination of, the criterion in cases where genuinely new information was being acquired. In such cases it might be that there is an indiscriminate predication of all attributes that could conceivably be applied to the *a*-term. In any event, whatever modifications are made to handle processing, the model will have to account for the fact that people usually identify anomalous statements very quickly (Glucksberg et al., 1982). This fact alone suggests that under normal circumstances, extensive searching for applicable attributes of the *b*-term without regard to their salience is improbable.

Whereas we have been concentrating on similarity there are other factors that contribute to understanding the nature of metaphors and metaphor comprehension. For example, Tourangeau and Sternberg (1981, 1982) provide a rather plausible account of metaphorical comprehension for some kinds of metaphorical statements (especially so-called genitive metaphors), but it is limited in its ability to account for the asymmetry phenomenon that was of interest in the studies reported here. Although Tourangeau and Sternberg (1982) are sympathetic to the salience model, believing that salience imbalance provides a "convincing" account of metaphorical asymmetry, other investigators, most notably Gentner (1980), have been more skeptical. In attempting to gather support for a

"structure-mapping" model (see Gentner, 1983), Gentner (1980) claims to have disconfirmed many of the main predictions of the imbalance model. In an experiment using eight metaphors, Gentner found that the interpretations of the statements (i.e., the grounds) were scored by judges as being predominantly "relational" rather than "attributional." She also found no salience difference of shared predicates for the two terms. Finally, she found no significant difference between forward and reversed metaphors with respect to either metaphoricity or aptness, although such differences as there were, were in the direction predicted by our asymmetry hypothesis.

In our view, there are problems with Gentner's study. First, her operational definition of salience in terms of order-of-mention of responses is, by her own admission, one that can be challenged--it confounds salience with ease of articulation. Thus, her failure to find salience imbalance may have been partly due to the fact that she did not have an adequate measure of salience in the first place. It may also have been due to the characteristics of the stimuli she used. All eight stimuli were selected from examples used in Ortony (1979a). Unfortunately, in many cases, they were examples designed to demonstrate potential problems and exceptions. For instance, some of Gentner's items (e.g., *encyclopedias are gold mines*) were used as illustrations in Ortony (1979a) of statements whose metaphoricity might be primarily due to factors other than salience imbalance (e.g., domain incongruence). Finally, the absence of a significant difference between forward and reversed items on either the metaphoricity or the aptness measures may have resulted from the fact that the items were not presented to subjects in *both orders*. Just as the spontaneous reversal phenomenon can mask differences in perceived similarity, so also might it mask differences with other measures. In this case also, her method of selecting items may have had unfortunate consequences. For example, the stimulus *surgeons are like butchers* and its reversed form *butchers are like surgeons*, were included in Ortony (1979a) to illustrate that some reversed metaphorical comparisons are interpretable in both directions, but with different grounds. For such items, the apparent comparability of metaphoricity and aptness ratings could be based on different readings.

Given the support for the salience imbalance hypothesis in our own studies, and those of others (e.g., Barsalou, 1982; Hanson, 1982; Katz, 1982), we are inclined to believe that salience imbalance is a common characteristic of metaphorical comparisons. This conclusion seems more defensible than the denial of the hypothesis based on establishing the null hypothesis with a small number of subjects and with a small number of (nonrepresentative) items. As pointed out in Ortony (1979a, p. 167) "many statements of similarity depend on some structural isomorphism between the knowledge associated with the two concepts rather than on merely a match of simple attributes." Gentner seems to share this view, but apparently considers it incompatible with the imbalance model.

Although we believe that salience imbalance properly defined is a major source of metaphoricity, it is surely not the only source. Nevertheless, while many unanswered questions remain, the present research does illuminate a number of issues that are of importance both to theories of similarity and of metaphor. The findings reported here suggest that the imbalance model is capable of accounting for the metaphoricity inherent in many similarity statements, as well as being able to provide a unified account of similarity judgments regardless of the kind of statements being judged.

References

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition, 10*, 82-93.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig and L. W. Porter (Eds.), *Annual Review of Psychology, 31*, 607-649.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283-319.

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*. New York: Academic Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Leaning and Verbal Behavior, 12*, 335-359.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.) *Discourse production and comprehension*. Norwood, NJ.: Ablex Publishing.

Cohen, L. J. (1979). The semantics of metaphor. In A. Ortony (Ed.), *Metaphor and thought*. New York: Cambridge University Press.

Gati, A., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 325-340.

Gentner, D. (1980). Studies of metaphor and complex analogies. Paper presented at the convention of the American Psychological Association.

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: the role of context. *Journal of Verbal Learning and Verbal Behavior, 22*, 577-590.

Glucksberg, S., Gildea, P. G., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior, 21,* 85-98.

Gregson, R. A. M. (1975). *Psychometrics of similarity.* New York: Academic Press.

Hanson, R. H. (1982). An investigation of the similarity and contrast models of metaphorical and categorical semantic processing. Unpublished doctoral dissertation. Purdue University.

Harwood, D. L., & Verbrugge, R. R. (1977). Metaphor and the asymmetry of similarity. Paper presented at the convention of the American Psychological Association.

Honeck, R. P., & Hoffman, R. R. (Eds.). (1980). *Cognition and figurative language.* Hillsdale, NJ.: Lawrence Erlbaum Associates.

Katz, A. N. (1982). Metaphoric relationships: The role of feature saliency. *Journal of Psycholinguistic Research, 11,* 283-296.

Katz, J. J., & Fodor, J. (1963). The structure of semantic theory. *Language, 39,* 170-210.

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology, 12,* 4-34.

Krumhansl, C. L. (1979). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 85,* 176-184.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Miall, D. S. (Ed.). (1982). *Metaphor: Problems and perspectives.* Atlantic Highlands, NJ.: Humanities Press, Inc.

Nygren, T. E., & Jones, L. E. (1977). Individual differences in perception and performance for political candidates. *Journal of Experimental Social Psychology, 13,* 182-197.

O'Hare, P. (Ed.). (1981). *Psychology and the arts.* Atlantic Highlands, NJ.: Humanities Press.

Ortony, A. (1979a). Beyond literal similarity. *Psychological Review, 86,* 161-180.

Ortony, A. (Ed.). (1979b). *Metaphor and thought.* New York: Cambridge University Press.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior, 17,* 465-477.

Rosch, E. & Mervis, C. D. (1975). Family resemblance studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Sacks, S. (Ed.). (1978). *On metaphor*. Chicago: The University of Chicago Press.

Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology, 7*, 82-138.

Smith, E. E. & Medin, D. L. (1981). *Categories and concepts*. Cambridge, Mass: Harvard University Press.

Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive Psychology, 13*, 27-55.

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11*, 203-244.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Tversky, A., & Gati, I. (1978). Studies in similarity. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 89*, 123-154.

Tversky, A., & Krantz, D. H. (1970). The dimensional representation and metric structure of similarity data. *Journal of Mathematical Psychology, 7*, 572-597.

Verbrugge, R. R., & McCarrell, N. S. (1977). Metaphoric comprehension: studies in reminding and resembling. *Cognitive Psychology, 9*, 494-533.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Footnote

The research reported here was supported in part by the National Science Foundation under Grant No. BNS-80-04541, in part by the National Institute of Education under Contract No. NIE-400-81-0030, and in part by a Spencer Fellowship awarded to the first author by the National Academy of Education. The second author is now with John C. Maloney & Associates, 620 North Michigan Avenue, Chicago, Illinois.

Special thanks are due to Stella Vosniadou for her many valuable contributions throughout the course of this research. We would also like to thank Robert R. Hoffman for his extensive comments on an earlier draft of this paper.

Table 1

Means and Standard Deviations for Literals and Similes on each Question

	Type of Question			
Item Type	Agreement	Metaphoricity	Adequacy	Correspondence
Literals	3.21 (.313)	1.58 (.267)	3.37 (.228)	3.22 (.316)
Similes	2.73 (.423)	2.97 (.345)	3.10 (.348)	2.73 (.517)
Control Literals	3.26 (.215)	1.43 (.089)	2.01 (.310)	1.77 (.280)
Control Similes	2.35 (.485)	2.72 (.417)	1.72 (.407)	1.37 (.301)

Table 2

Means and Standard Deviations for

Literals and Similes on each Question

	Type of Term	
Item Type	*a*-term	*b*-term
Literals	4.21	4.10
Similes	3.52	3.81

Appendix A

Literals

an anvil is like an anchor
an art gallery is like a museum
a scalpel is like a razor
a visa is like a passport
an anesthetic is like a sleeping pill
a lamp is like a light
a game reserve is like a zoo
an oil well is like a gold mine
an eel is like a snake
a crab shell is like a tortoise shell
a steamroller is like a bulldozer
a hand grenade is like a time bomb
an escalator is like a stairway
a stream is like a river
a scavenger is like a parasite
an ant hill is like a beehive
a landslide is like an earthquake
a horseshoe is like a magnet

Anomalies

a cigarette is like a school
a mountain road is like a visa
a scalpel is like a horseshoe
a stream is like a hand grenade
a game reserve is like a sermon
an oil well is like a sleeping pill
a steamroller is like a gold mine
a razor is like an eel
an art gallery is like a tortoise shell
a visa is like a tranquilizer
an iceberg is like a passport
an anvil is like a beehive
a polite manner is like an escalator
an ant hill is like an earthquake
a library is like a suburb
a glacier is like an artery
a tyrant is like a river
a time bomb is like a smile

Similes

an education is like a stairway
a political revolution is like an earthquake
an artery is like a river
a mountain road is like a snake
a sermon is like a sleeping pill
a library is like a gold mine
a friend is like an anchor
a cigarette is like a time bomb
a mobile home is like a tortoise shell
a family album is like a museum
a book is like a light
a sarcastic remark is like a razor
a smile is like a magnet
a university is like a beehive
a suburb is like a parasite
a tyrant is like a bulldozer
a school is like a zoo
a polite manner is like a passport

Appendix B

Note: Mean metaphoricity ratings appear in parentheses following each comparison and mean adequacy ratings follow each of the grounds.

Similes and Grounds (List 1)

illiteracy is like a prison (3.00)
inhibiting freedom and restricting opportunity (3.16)
schools are like zoos (2.91)
being disorganized and noisy (2.83)
cities are like beehives (3.25)
being center of work and activity (3.42)
slums are like tumors (2.91)
being ugly and growing uncontrollably (3.00)
desks are like junk yards (2.75)
being disorganized and littered with things no longer wanted (2.83)
soldiers are like pawns (3.25)
being expendable and of relatively little value (2.75)
children are like snowflakes (2.91)
being pure and individually unique (3.00)
reality is like a sledgehammer (3.08)
delivering heavy and sudden blows (3.08)
salesmen are like bulldozers (3.00)
being relentless and undeterred by obstacles (3.50)
drug pushers are like lepers (2.66)
being dangerous and spurned by society (2.75)
subway systems are like mazes (3.08)
being structurally complicated and easy to get lost in (3.50)
an education is like a ladder (3.41)
being a means of reaching higher places (3.42)
clouds are like cotton balls (3.25)
being white and fluffy in appearance (3.42)
a friend is like an anchor (3.08)
providing stability during turbulent times (3.33)
cats are like princesses (2.58)
being fussy and expecting special treatment (2.75)
insults are like razors (3.25)
cutting deeply and causing pain (3.58)
family albums are like museums (3.08)
containing records of the past (3.41)
smiles are like magnets (3.17)
attracting things in their vicinity (3.08)
roads are like snakes (3.33)
being twisted in appearance and unpredictable (3.08)
rage is like a volcano (3.58)
erupting unexpectedly and violently (3.75)

Literals and Grounds (List 1)

lectures are like sermons (2.00)
imparting information orally to groups of people (3.16)
landslides are like earthquakes (1.33)
being catastrophic events involving the landscape (3.58)
moths are like butterflies (1.33)
coming from cocoons and flying (3.58)
switchblades are like daggers (1.66)
being sharp weapons that are easily concealed (3.25)
a mobile home is like a trailer (1.33)
being lived in and movable to new locations (3.58)
libraries are like archives (1.91)
storing and preserving knowledge (3.58)
ghouls are like vampires (1.58)
being scary creatures featured in books and movies (3.50)
obligations are like responsibilities (1.75)
being moral commitments that should be honored (3.75)
tire irons are like crow bars (1.50)
being levers made of metal (3.08)
hair is like fur (1.66)
growing on skin and providing warmth (3.17)
paste is like glue (1.50)
being a sticky substance sold in bottles (3.08)
faith is like belief (1.83)
being a state of mind involving acceptance (3.00)
idols are like gods (1.91)
being revered by those who believe in them (3.33)
cigarettes are like cigars (1.41)
being cylindrical and made of tobacco (3.33)
proverbs are like maxims (1.75)
being wise sayings considered to be true (2.92)
a visa is like a passport (1.41)
being a document needed for foreign travel (3.50)
shopping centers are like markets (1.58)
being places where goods are bought and sold (3.50)
high rises are like skyscrapers (1.58)
being tall buildings usually situated in cities (3.50)
intermissions are like time-outs (2.00)
being temporary breaks in entertainment (3.17)
televisions are like radios (1.58)
being electronic devices that present information (3.33)

Control Similes with Inadequate Grounds (List 1)

fog is like a coat (3.08)
making mobility and running difficult (1.91)
fugitives are like foxes (2.75)
having teeth and ears (1.16)
hallways are like telescopes (2.50)
being nice to own and use (1.00)

perfume is like a tool (2.33)
being purchased in specialized stores (1.83)
stagecoaches are like dinosaurs (2.50)
being discussed in books about the past (1.91)
rumors are like viruses (3.50)
seldom changing or going away by themselves (2.66)

Control Literals and Inadequate Grounds (List 1)

groves are like forests (1.41)
being good places for picnics (2.16)
jobs are like careers (1.41)
being sources of pleasure and disappointment (2.16)
rivers are like streams (1.50)
having beginnings and ends (1.74)
ponds are like lakes (1.58)
containing living organisms and varying in temperature (2.41)
roosters are like hens (1.41)
having small bodies and legs (1.83)
refrigerators are like iceboxes (1.33)
being big and difficult to move (1.25)

Similes and Grounds (List 2)

lectures are like lullabies (2.66)
inducing drowsiness and sleep (3.25)
political revolutions are like earthquakes (2.91)
involving upheaval and danger (3.00)
dancers are like butterflies (3.83)
being graceful and beautiful (3.50)
lies are like daggers (3.25)
being intended to wound and not used openly (2.83)
a mobile home is like a turtle shell (3.00)
being movable and just large enough for their inhabitants (3.00)
libraries are like gold mines (3.58)
being sources of wealth and worthy of exploration (3.25)
debt collectors are like vampires (3.25)
taking the necessities of like from unwilling victims (3.00)
obligations are like shackles (3.08)
restricting behavior and choice (3.16)
questions are like crow bars (2.91)
being used for prying out and extracting things (3.42)
hair is like spaghetti (2.83)
having tangled and flexible strands (3.33)
trust is like glue (3.33)
producing a strong and permanent bond (3.50)
faith is like a beacon (2.50)
providing direction and guidance (2.83)
surgeons are like gods (2.75)
being of unquestioned authority and having the power over life and death (2.75)

cigarettes are like pacifiers (3.16)
providing oral satisfaction and soothing (3.41)
proverbs are like spotlights (2.58)
focusing on and illuminating things of importance (3.08)
glamor is like a passport (2.83)
facilitating access to desired places (3.33)
shopping center are like jungles (3.50)
being complex and difficult to find one's way around in (3.41)
giraffes are like skyscrapers (3.16)
being much taller than other things of their kind (2.92)
vacations are like time-outs (2.50)
being opportunities to rest and revitalize (3.58)
busy-bodies are like radios (2.66)
broadcasting information indiscriminately (3.16)

Literals and Grounds (List 2)

a reform school is like a prison (1.91)
holding law-breakers in a secure environment (3.33)
farms are like zoos (1.25)
being places where animals are cared for (3.75)
cities are like towns (1.33)
being places with buildings inhabited by groups of people (3.00)
slums are like ghettos (1.33)
being run-down urban areas (3.50)
garbage dumps are like junk yards (1.83)
being places for disposing of unwanted objects (3.58)
soldiers are like sailors (1.50)
wearing uniforms and being members of the armed forces (3.42)
raindrops are like snowflakes (1.33)
being individual units of precipitation (3.50)
a mallet is like a sledgehammer (1.25)
having a strong handle and being used for driving stakes (3.66)
steamrollers are like bulldozers (1.33)
being heavy machines used for flattening the ground (3.58)
drug pushers are like pimps (1.66)
being involved in illegal business enterprises (3.08)
subway systems are like railroad systems (1.16)
being networks of routes for trains (3.75)
a stairway is like a ladder (1.58)
being used for ascending and descending (3.50)
clouds are like mists (2.16)
being masses of air saturated with water (3.25)
a friend is like a relative (1.91)
being a person relied upon in emergencies (2.83)
princes are like princesses (1.33)
being royality in line for the throne (3.50)
scalpels are like razors (1.41)
being very sharp and used in precise cutting (3.58)
art galleries are like museums (1.66)
being public buildings containing valuable artifacts (3.25)

smiles are like grins (1.33)
being signs of amusement and pleasure (3.41)
roads are like paths (1.58)
being land routes from one place to another (3.50)
rage is like fury (1.50)
being an emotion that is difficult to hide (3.08)

Control Similes with Inadequate Grounds (List 2)

beards are like forests (2.25)
growing and being found all over the world (1.83)
jobs are like jails (2.16)
being places with people and equipment (1.83)
rivers are like ribbons (2.75)
being used and admired (1.33)
ponds are like mirrors (2.91)
having edges and flat surfaces (1.74)
roosters are like clocks (3.41)
having faces an moving parts (3.00)
marriages are like iceboxes (2.50)
losing their novelty quickly (1.66)

Control Literals with Inadequate Grounds (List 2)

a jacket is like a coat (1.25)
being convenient and lasting for many years (1.75)
wolves are like foxes (1.50)
having legs and breathing (1.91)
microscopes are like telescopes (1.50)
being built and used by technicians (2.16)
perfume is like cologne (1.33)
being a liquid dispensed in bottles (2.25)
stagecoaches are like wagons (1.41)
being made of wood and metal (2.25)
germs are like viruses (1.50)
being small and found everywhere (2.24)

Appendix C

Similes (Set 1)

ballerinas are like butterflies
slums are like tumors
reality is like a sledgehammer
an education is like a ladder
a friend is like an anchor
insults are like razors
smiles are like magnets
illiteracy is like a prison

Similes (Set 2)

encyclopedias are like gold mines
trust is like glue
cities are like beehives
rage is like a volcano
obligations are like shackles
sermons are like sleeping pills
shopping centers are like jungles
debt collectors are like vampires

Literals (Set 1)

diplomas are like certificates
streams are like rivers
raindrops are like snowflakes
cigarettes are like cigars
garbage dumps are like junk yards
sores are like warts
landslides are like earthquakes
steamrollers are like bulldozers

Literals (Set 2)

stagecoaches are like wagons
germs are like viruses
ponds are like lakes
faith is like belief
beards are like moustaches
televisions are like radios
mobile homes are like trailers
perfume is like cologne